Soft sensor development: mining process quality prediction

Contents:

---

1. Dataset description

The dataset originates from a mining process, from the froth flotation phase. The process is successful if there is as little outlet silica content as possible. There are 24 variables (columns) in the data. They are presented in the table below.

Variable no.	Variable name	Variable no.	Variable name
1	Date	13	Flotation column 5 air flow
2	Iron conc. In feed [%]	14	Flotation column 6 air flow
3	Silica conc. In feed [%]	15	Flotation column 7 air flow
4	Starch flow	16	Flotation 1 column level
5	Amina flow	17	Flotation 2 column level
6	Ore pulp flow	18	Flotation 3 column level
7	Ore pulp pH	19	Flotation 4 column level
8	Ore pulp density	20	Flotation 5 column level
9	Flotation column 1 air flow	21	Flotation 6 column level
10	Flotation column 2 air flow	22	Flotation 7 column level
11	Flotation column 3 air flow	23	Outlet iron conc. [%]
12	Flotation column 4 air flow	24	Outlet silica conc. [%]

2. Modelling goals

a. Level B (20p) Silica content soft sensor

Predict the outlet silica content in the ore. Investigate the most important variables in prediction, and how they vary with the silica content, using model weights and regression coefficients. Evaluate your model prediction.

***Note**: Since it is a B-level task, you can use the outlet iron concentration as an input variable to the model.*

b. Level A (30p) Dynamic soft sensor for silica content

Predict the outlet silica in the ore using dynamic models. Investigate how the regression coefficients change with time. Use lagged variables and reduce the variables to the minimum possible. Determine the best window frame for accurate predictions.
***Note**: Since it is a A-level task, you cannot use the outlet iron concentration as an input variable to the model (unless it is a lagged variable).*

c. **Level A2 (30p)** Future predictions of silica content

Using only lagged variables and past known values of silica ore, calibrate a model that can estimate the future values of silica content in the ore. You can use dynamic models. Evaluate the most important variable for prediction.
***Note**: Since it is a A-level task, you cannot use the outlet iron concentration as an input variable to the model (unless it is a lagged variable).*

d. Level A3 (30p) Global model with optimal lag

Passing through the process, a batch is at different times in different units. Determine the optimal lag for each variable, that will make prediction accurate for a global model. You are allowed to use lagged values of the output variable as an input (and lagged values of iron ore).
***Note**: Since it is a A-level task, you cannot use the outlet iron concentration as an input variable to the model (unless it is a lagged variable).*

---

3. **Dataset hints**

- The dataset is known to contain outliers or abnormal functioning periods of the process. It is up to you if you choose a chunk of data in which the process is normal, and another chunk for testing your model.
- If you choose to simply exclude the periods of abnormal functioning, take care when you lag the variables ☺. You don't want to end up with the flotation level of last hour being the one 24h ago!
- The variables are not captured with the same frequency. Each observation has data for each column, but some observations don't change in a 60-minute frame. You can decide if you want to resample the whole dataset to have the same sampling rate (by averaging the other observations that have more frequent observations - *downsampling*, or run a filter through the ones that have less observations - *upsampling*).

---

4. References

[1] - Dataset: Quality Prediction in a Mining Process [accessed online at: https://www.kaggle.com/edumagalhaes/quality-prediction-in-a-mining-process, last accessed: 17.08.2023]

[2] - Zhang, Y., Teng, Y. & Zhang, Y., 2010. Complex process quality prediction using modified kernel partial least squares. Chemical Engineering Science, 65(6), pp.2153–2158.
[3] - Wang, D., Jun Liu & Srinivasan, R., 2010. Data-Driven soft SENSOR approach for Quality prediction in a refining process. IEEE Transactions on Industrial Informatics, 6(1), pp.11–17.
[4] - Ge, Z. et al., 2014. Two-level pls model for Quality prediction Of multiphase batch processes. Chemometrics and Intelligent Laboratory Systems, 130, pp.29–36.